ACARM *Newsletter*, Issue 36, Summer 2005

Electronic Reference Media in The National Archives of India: An Overview

N.S. Mani, Microphotographist, National Archives of India

The National Archives of India (NAI) is the custodian of the non-current records of the Government of India and holds them in trust for the use of administrators and scholars. It is an Attached Office of the Ministry of Culture Government of India. It was set up in March 1891 in Calcutta (Kolkata) as the Imperial Record Department and subsequent to the transfer of the National Capital from Calcutta to New Delhi in 1911 it shifted to its present building in 1926. The nomenclature of the Imperial Record Department was changed to National Archives of India after India gained independence on 15th August 1947.

The NAI has in its custody records occupying 30 kilometres of linear shelf space. Most of these records relate to the East India Company and the British Raj and are in English. It has also acquired the private papers of eminent Indians including Mahatma Gandhi, Rajendra Prasad, Maulana Azad, Sardar Patel, M.R. Jayakar, Field Marshal K.M. Cariappa and others. Approximately 1,73,000 books comprising rare Government reports, proscribed publications, vernacular Native Newspaper Reports, Parliamentary Papers, etc. are available in the Library of the National Archives. The Department also has in its custody more than 7000 microfilms of records of Indian interest acquired from other countries. As a long-term preservation measure, the NAI has been engaged in security microfilming of important records series and approximately 12,000 microfilms of public records have been prepared.

The last decade of the 20th century witnessed the gradual globalization of archives due to the rapid pace of technological progress and thus necessitated the need to evolve a method for facilitating the quick retrieval of information. Computerization of archives, which appeared to be the best solution, was then in its infancy, consequently the NAI initiated steps for the development of software for archival use, as there was no internationally recognized and standard software available for archival purposes. At that time most of the countries were laying down specifications for hardware and software according to the nature of the record holdings in their custody. They developed software for Archives in Lotus, dbase-III, FoxBASE, Word Perfect under DOS/Win-3.1 and UNIX environments as information stored was mostly in textual format (ASCII and flat file).

By 1998, with the further advancement of information technology and the availability of new storage mediums, the NAI considered the question of the creation of a data base for reference media records to enable the quick retrieval of information as a viable alternative to save time. It was considered to be economical for the research and reference purposes of scholars and other users of archives. However, many issues were involved and needed to be resolved before the creation of the database. As the storage of electronic documents is vulnerable to loss by decay of the medium in which information is stored, they become inaccessible and unreadable especially when the software needed to interpret them or the hardware on which the software runs becomes obsolete. The advancement in information technology requires the frequent migration of data to the latest medium. Most digital documents exist only in encoded form, which require specific software to translate the bit streams and make them usable. The medium itself disappears from the market; drivers capable of controlling the drives and deciphering the encoding used on the medium, are no longer incorporated in new computers. Upgrading of new computer systems, therefore, requires abandoning the old storage medium, provided an institution still has in its custody documents stored on that medium. The problems of short lifespan of the media and rapid obsolescence have necessitated copying of the digital information onto new media. Although copying is a viable solution, it is likely to corrupt documents due to compression, encryption or ever changing data formats.

Considering all these aspects, the NAI's first priority was to computerize data of reference media to records for easy retrieval and access for the benefit of scholars and other users of archives. Accordingly a tailor-made software "AIMS" (Archival Information Management System) was developed by NAI in Visual Basic back end and MS Office front end. With the introduction of this software, the work of data entry is progressing satisfactorily.

The data entry is now being created from the following formats:

- Index to press lists, press lists, and transfer lists
- Printed (type written) calligraphy form (hand written) and manuscripts in various sizes and styles
- Languages - Hindi, English, Persian, Urdu and Sanskrit

This was, however not an easy task due to a number of problems faced, *viz*, diverse forms and formats of records; no fixed data structure; cross linking of records; no existing standards for computerization; multiple keywords within the subject field.

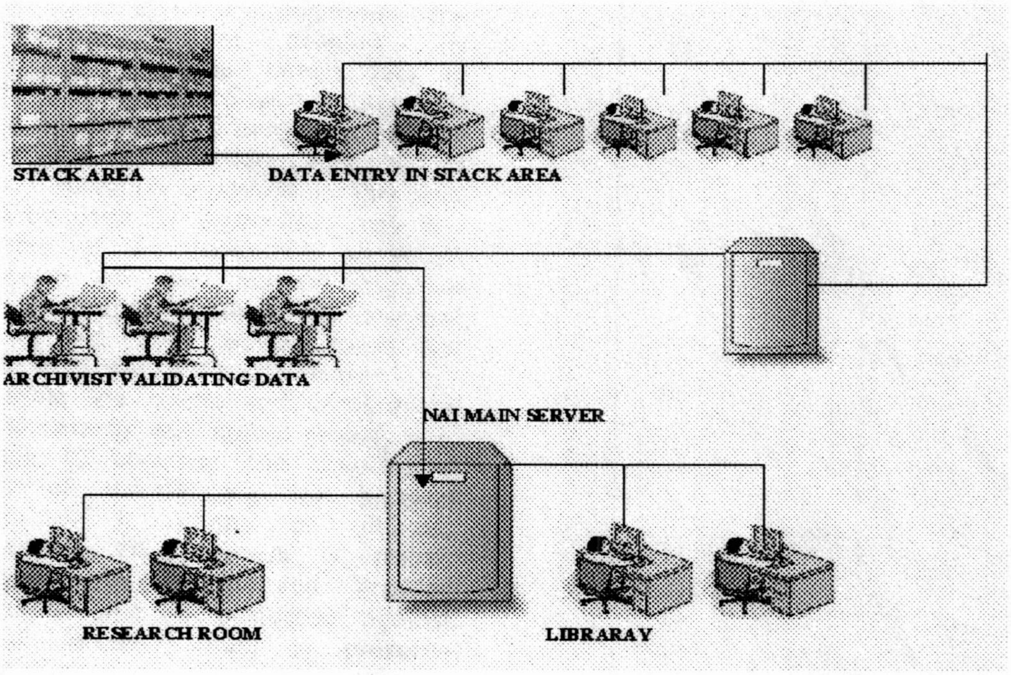
Simultaneously, the NAI designed an integrated system solution for the Archival Information and Management System (AIMS) software on intranet ware. The software was developed with web browser based access and having facility to upgrade to web/intranet and WAN deployment. The software had a flexible modular design and a package was developed to facilitate computerization of indexes to records. Also, AIMS software was customized to access the information related to an important historical event, chronological span, and activities of a personality, political leaders and much more.

The AIMS software has an integrated system – LAN implementation, six modules and ten sub modules and three levels of security, online processing and retrieval besides report generation and web/internet compatibility. The software was developed with a SQL search engine to locate the file details on the basis of catchwords, Records browsing in Departments, Branch, Year, Month and File Details, along with subject and keywords as well as tree and flat view retrieval.

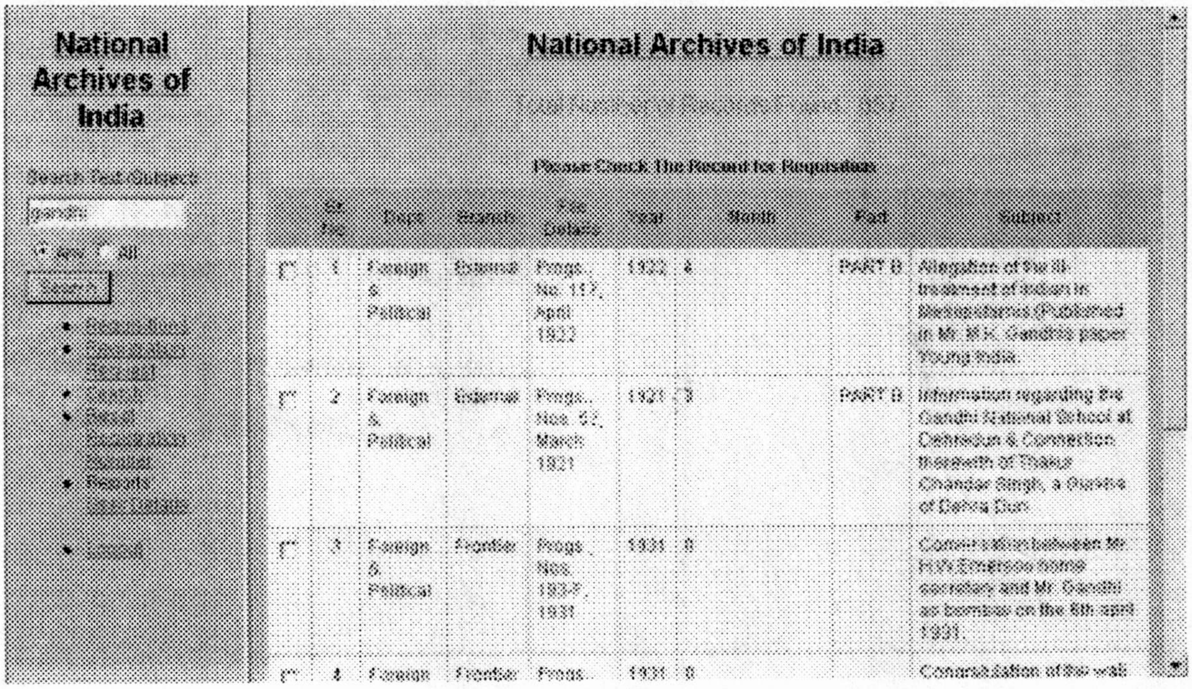
There are around 52,00,000 lakhs records for which reference media are to be created. So far reference media has been created for 12,00,000 records and are available for retrieval through AIMS software primarily for the benefit of scholars and users of archives on the NAI intranet. A database of the entire collection of the NAI Library has also been created in LIBSYS software and is available for retrieval on the NAI intranet.

Since the Department has a huge collection of microfilms of documents available for use by scholars, a pilot project to convert the analogue microfilm images to digital medium was undertaken in 2000. Around 2,000,000 lakhs images have been created and the AIMS software is being updated in order to make available digital images of the documents on the network, so that frequent use of any document is minimized and multiple viewers are able to have easy access to the original files in digital form on the NAI intranet.

Graphic Representation of Data Entry and Data Validation for Reference Media of Records in National Archives of India, New Delhi.



Retrieval of information on catch word Gandhi in AIMS software



gandhi

● Any ○ All

National Archives of India

Please Check The Record for Requisition

	Sr. No.	Dept	Branch	File Details	Year	Month	Part	Subject
☐	1	Foreign & Political	External	Progs., No. 117, April 1922	1922	4	PART B	Allegation of the ill-treatment of Indians in Mesopotamia (Published in Mr. M.K. Gandhi's paper Young India.
☐	2	Foreign & Political	External	Progs., Nos. 52, March 1921	1921	3	PART B	Information regarding the Gandhi National School at Dehradun & Connection therewith of Thakur Chandar Singh, a Gurkha of Dehra Dun.
☐	3	Foreign & Political	Frontier	Progs., Nos. 193-F, 1931	1931	8		Conversation between Mr. H.W. Emerson home secretary and Mr. Gandhi at bombay on the 8th april 1931.
☐	4	Foreign	Frontier	Progs.	1931	8		Congratulation of the wali

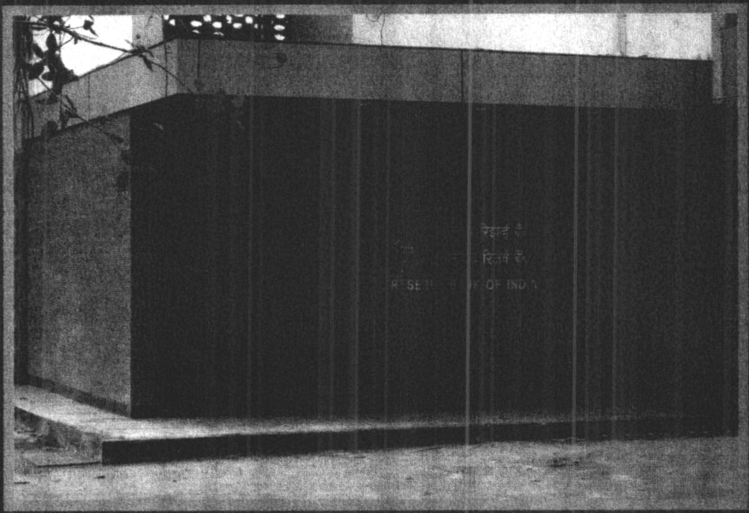
It is pertinent to mention that even in countries where e-records or digital documents are created and transferred to their national archival repositories, the records are "future proofed" by transferring the digital image data to microfilm. However, converting digitized data to microfilm as a long-term storage medium involves a notable reduction in quality. Microfilm produced in this way cannot be used for digitization with any guarantee of an acceptable result in comparison to the master negative prepared from original documents. Analogue and digital storage forms are thus not fully compatible at present.

Considering the merits and demerits of both microforms and digital media, it is advisable to continue the time-tested process of microfilming for archival records as a means of ultimate preservation. This process is stable, interfaces with digital and analogy technologies and is protected against computer viruses. Thus, to preserve archival documents, and avoiding the risk of content alteration, for indeterminate periods into future, a system that balances the benefits of microfilm media with adequate search facilities and organizational aspects of a digital system with proper indexing for rapid retrieval and online access, appears to be an ideal solution for an archival institution. Until such time as a viable and standardized solution for digital longevity is developed and accepted as a long term solution for preservation of record data for posterity, we have to content ourselves with the traditional medium of preserving by microforms and allied reprographic techniques.

For more information contact: *N.S.Mani, National Archives of India, Janpath, New Delhi, Email*: archives@ren02.nic.in. ◆

25

1981-2006

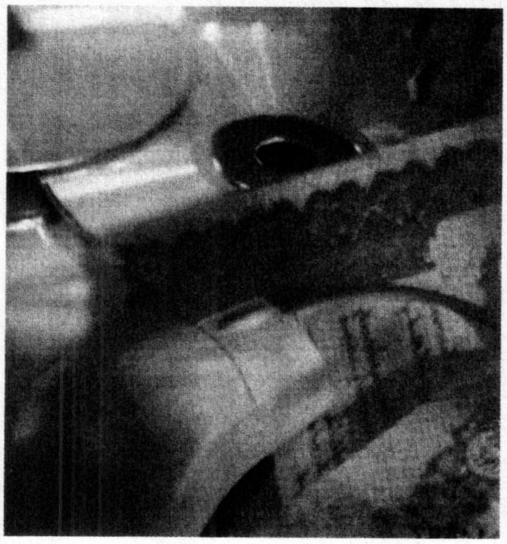


स्मारिका
Souvenir

भारतीय रिज़र्व बैंक अभिलेखागार
RESERVE BANK OF INDIA ARCHIVES

Recent Trends in the Digitalisation of Records

N.S. Mani



Retrieval of information from paper heritage is a difficult and time consuming task. The digitisation of documentary heritage is fast emerging as an important tool for solving difficulties of retrieval in so far as consultation and dissemination of information incorporated in the documents are concerned. Digitalisation is a modern process in which images on paper are transformed into electronic format, which could be shared by all through worldwide network, thus enabling simultaneous sharing of information by number of users.

Document scanners, micrographic equipment, viz. microfilm scanner and digital cameras are important array of equipment, which are universally in vogue to decode document images. When the document is scanned, the image is converted into TIFF or JPEG format, which provides a quality archival image. This also retains the visual formatting and layout of the document intact. Electronic records could be accessed by world wide web and instantaneously copied or printed, thereby eliminating the time and cost involved to transfer them even to the remotest corner of the world. The advantages of electronic images are a boon and serves as a convenient mechanism to locate a document at a given work place.

Why Digitisation?

1. To increase access.
2. To reduce dependence on handling fragile and brittle documents besides original documents.
3. To ensure and obtain a back up medium / copy.

The National Archives of India (NAI) initiated a pilot project for digitisation of rare manuscripts with a skeletal unit comprising Sony Digital Camera (2.1 mega pixel) and a designated personal computer in July 2000. This Unit was set up in collaboration with the Department of Science and Technology and the National Institute of Advanced Studies, Bangalore. Rare manuscripts, viz. Bhagwadgita, Ramayana, Mahabharata, etc. were digitised and stored in CD medium. A total of 6000 images have since been created. Simultaneously a pilot project was undertaken to convert the analog microfilm images of documents to digital medium. In pursuance of this project 2,00,000 images have so far been created. Apart from this Archival Information Management System (AIMS) software is being used in the National Archives of India for retrieval of information from the available reference media of records. This software is being updated with a view to facilitate display of digital images of documents on the monitors on the network. This would greatly reduce the dependence on the frequent use of documents and also facilitate multiple access to the original files in the National Archives of India in digital format on the intranet. Once this project is implemented in the LAN of NAI, the same would be hosted on a server for wider access with facility for making payment online for obtaining copies of the requisite pages from anywhere.

Records Management techniques all over the world are undergoing a sea change. Naturally, this calls for reassessing, rethinking and redesigning almost every aspect of traditional records format, which is presently in vogue. In order to have a better understanding and to meet the challenges posed by this revolutionary aspect, every Archives has to adopt new methodology for the management of its records in a fresh perspective.

However, the following should be kept in mind while defining the record management policy:

1. Policy regarding what should be digitalised.
2. Procedure for implementing the policy.
3. Migration system for accommodating and adopting the frequent changes in technologies.
4. Document images storage systems, which should have ability for adaptation to fast changing technologies.

We are in the 21st century where we are witnessing large scale generation of electronic records due to government's thrust in the field of information technology and it is a universal fact that the digital records whether paper records are vulnerable to various risks and dangers due to their corruption and eventual loss. However, under the stipulations contained in the Public Records Act, 1993 and coming into effect of the Public Records Rules, 1997, it is imperative that records should not be transferred to the Archives until they are 25 years old and be preserved in the Departmental Record Room. However, with the introduction of computers in the Government of India offices, large scale records are being generated in electronic format along with generation in paper format for storing the electronic files in hard disks and CD-ROMs for ensuring preservation in standby mode. It is, therefore, essential that the National Archives of India should embark upon a programme for selection and preservation of electronic data during the stage of creation itself.

Electronic records, which are being created in the Government offices, are not being transferred to the National Archives of India and are being presently stored in the hard disk and in CD medium by the creating agency. Action should be taken to transfer the digital records to the National Archives of India after their appraisal by the respective creating agencies. The digital data should be stored in a proper medium for long term storage under ambient conditions. Although the process of preservation of digital data is labour intensive, no preservation activities have been initiated at present as the records are in the stage of infancy. The ideal way to manage the burgeoning electronic record are through (i) identification, (ii) selection, and (iii) elimination, which should be initiated by the creating agency as soon as they are created.

The storage mediums for the digital images are normally the hard disk, CD- ROM or magnetic tapes. The claims made by experts with regard to the longevity of these digital mediums are not time tested and their life span ranges from ten to fifteen years depending on the quality of CD. The silver reflective layer usually gets tarnished due to light, heat, humidity, oxygen and also with rough handling which results in loss of data, which would otherwise had a longevity of 25 years. In order to offset this problem, Kodak Company had in 2006 introduced a 24 Karat Gold CD ROM and DVD's which they claim has extended life span and the longevity of stored data to many years. This Company has claimed that with the launch of gold layered CD ROM, the storage of digital data for 300 years would become a virtual reality. Besides, with the introduction of gold layered DVD-R, one can ensure protection of data in videos and digital files for a period ranging from 80-100 years.

The 24 Karat gold reflective layer and CD-R and DVDs would ensure protection of the medium of storage and would certainly not tarnish or oxidize the virtual data stored in the CD-ROM or DVDs. These new CDs are coated with Phthalocyanine dye, which is an extremely stable medium for ensuring longevity of the primary data storage layer on the CD-R. Besides, a triple layer of protective medium on the CD-R ensures protection against heat, humidity, bright light and the ill effects of rough handling. The substrata of CD-R are usually made of optical grade polycarbonate, which is better than many CDs available as claimed by Kodak Company.

With the rapid advancement in the digital technology, many issues continue to remain unresolved. Digitalisation is an expensive affair, which requires detailed planning for ensuring continual access to digital data. It is pertinent to mention that digitilisation is not preservation, not cheaper, safer or more reliable than microfilming programme.

Digitalisation Programme is therefore, not to be recommended for converting "microfilm programme" to "digitization programme" as it has inherent risks involved in ensuring long term preservation. However, it could serve as a component for protection of certain documents. If an organization has an ultimate aim to save space as a corollary to preservation of documents, it would certainly be advisable to create microfilm copies rather than digital images.

The juxtaposition or confluence of microfilm and digital technologies has been inseparable feature of information technology for the last 30 years. Microfilm and digital technologies could be effectively combined to meet the need of both archival storage and digital access. This hybrid technology enables microfilm and scanning simultaneously with a single click with effective results one for permanent storage and the other for online access. Till such time, capabilities of digital preservation have been proved as being cost effective, microfilming would remain the primary reformatting strategy for ensuring records preservation. Microfilm offers acceptable levels of quality media, longevity, and reduces dependence on machines and serves as viable means for producing number of copies with no loss of information contents.

The Archives in India, both at national and state level, should have the will and ability to preserve records by frequently migrating electronic records to new format. This would ensure insulation against obsolescence. In the present scenario, Electronic Records Management (ERM) appears to be the domain of IT managers who hardly give credence to preservation of electronic data and give it the attention and priority to the current requirements. It is felt that the preservation needs also receive equal attention and importance in any records management programme followed / adopted by the archival institutions.

■

Electronic Reference Media in National Archives of India – An Overview

The National Archives of India is the custodian of the non-current records of the Government of India and is holding in trust for the use of administrators and scholars. It is an Attached Office of the Ministry of Culture Government of India. It was set up in March 1891 in Calcutta (Kolkata) as the Imperial Record Department and subsequent to the transfer of the National Capital from Calcutta to New Delhi in 1911 it shifted to its present building in 1926. The nomenclature of the Imperial Record Department was changed to National Archives of India after India gained independence on 15th August 1947.

The National Archives of India has in its custody records occupying 30 kilometers of linear shelf space. Most of these records relate to East India Company and the British Raj and are in English. It has also acquired private papers of eminent Indians including Mahatma Gandhi, Rajendra Prasad, Maulana Azad, Sardar Patel, M.R. Jayakar, Field Marshal K.M. Cariappa and others. Approximately 1,73,000 books comprising rare Government reports, proscribed publications, vernacular Native Newspaper Reports, Parliamentary Papers, etc. are as well available in the Library of the National Archives. The Department also has in its custody more than 7000 microfilms of records of Indian interest acquired from other countries. As a long-term preservation measure, National Archives of India has been engaged in security microfilming of important records series and approximately 12,000 microfilms of public records have since been prepared.

The last decade of the 20th century witnessed gradual globalization of archives due to the rapid pace of technological progress and thus necessitated the need to involve the method for facilitating quick retrieval of information. Computerization of archives, which appeared to be the best solution was, then in its infancy, consequently the National Archives of India initiated steps for the development of software for archival use, as there was no internationally recognized and standard software available for archival purposes. At that time most of the countries were laying down specifications for hardware and software according to the nature of their record holdings, which were in their custody. They developed software for Archives in Lotus, dbase-III, FoxBASE, Word Perfect under DOS/Win-3.1/and UNIX Environment as information stored was mostly in textual format (ASCII and flat file).

By 1998, with the further advancement of information technology and availability of new storage medium, the question of creation of data base for reference media of records for quick retrieval of information from the printed reference media was being considered in the National Archives of India as a viable alternative to save time. It was considered economical for research and reference purpose for scholars and users of archives. However many issues were involved and needed to be solved before creation of the database of the reference media of records available in archives. As the storage medium of electronic documents were vulnerable to loss by decay of the medium in which information is stored, they became inaccessible and unreadable especially when the software needed to interpret them or the hardware on which the software runs became obsolete. The advancement in information technology required migration of data to the latest medium frequently. Most digital documents exist only in encoded form, which required specific software to bring their bit streams to action and make them usable. The medium itself disappears from the market; drivers capable of controlling the drives and deciphering the encoding used on the medium, are no longer incorporated in new computers. Upgrading of new computer systems, therefore, requires abandoning the old storage medium, provided an institution still has in its custody documents stored on that medium. The problems of short lifespan of the media and rapid obsolescence have necessitated copying of the digital information into new medium. Though copying is a viable solution, it is likely to corrupt documents due to compression, encryption or ever changing data formats.

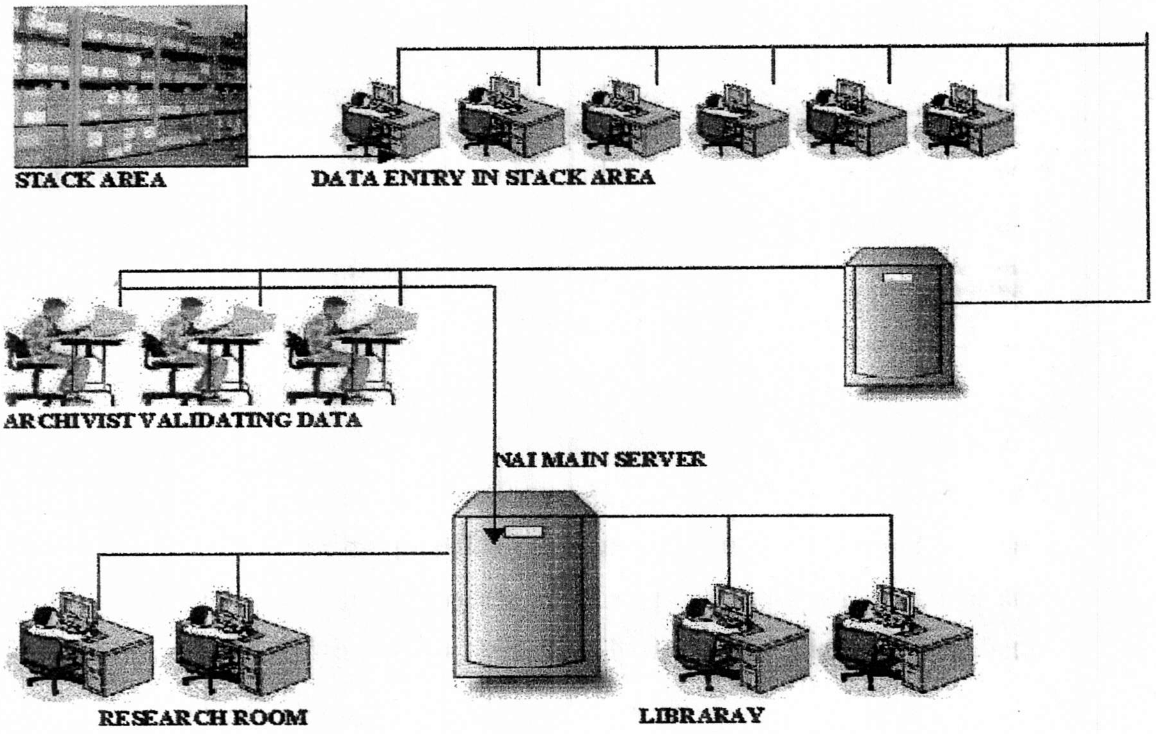
Considering all these aspects, the National Archives of India's first priority was to computerize data of reference media to records for easy retrieval and access for the benefit of scholars and users of archives. Accordingly a tailor made software "AIMS" (Archival Information Management System) was developed by National Archives of India in visual basic front end and MS Office back end. With the introduction of this software, the work of data entry is progressing satisfactorily.

The data entry is now being created from the following formats.

1. Index to press lists, press lists, and transfer lists
2. Printed (type written) calligraphy form (hand written) and manuscripts in various sizes and styles
3. Language - Hindi, English, Persian, Urdu and Sanskrit

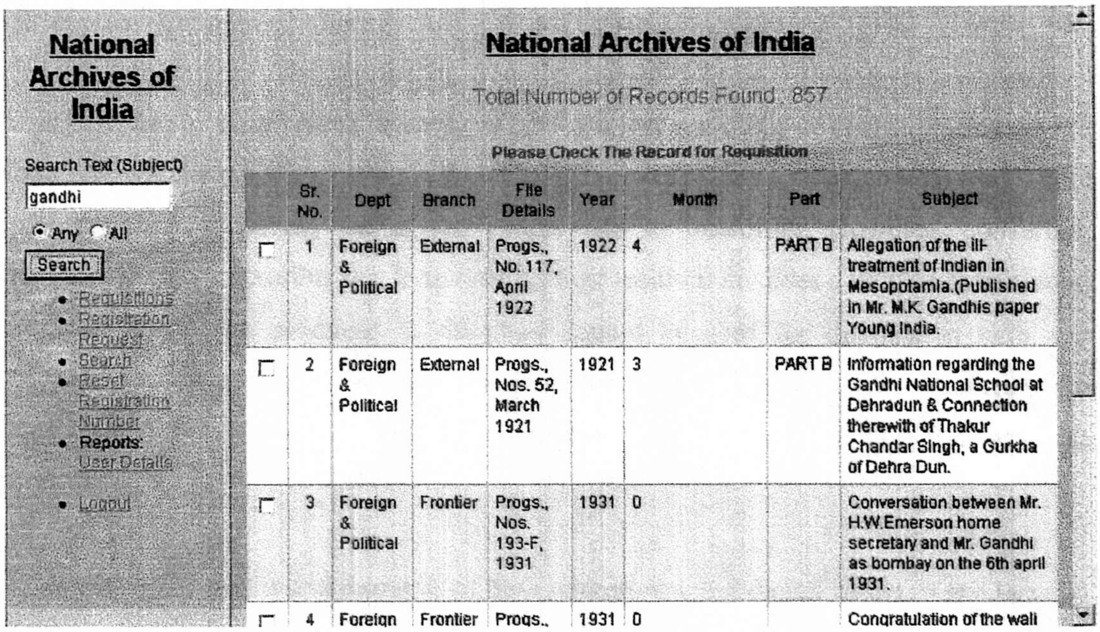
This was, however not an easy task due to a number of problems faced, viz., diverse forms and formats of records; no fix data structure; cross linking of records; no existing standards for computerization; multiple keywords within the subject field.

Graphic Representation of Data Entry and Data Validation for Reference Media of Records in National Archives of India, New Delhi.



Simultaneously, the National Archives of India designed an integrated system solution for the Archival Information and Management System (AIMS) software on intranet ware. The software was developed with web browser based access and having facility to upgrade for web/intranet and WAN deployment. The software had flexible modular design and a package was developed to facilitate computerization of indexes to records. Besides, AIMS software was customized to access the information related to an important historical event, chronological span, and activity of a personality, political leaders and much more.

Retrieval of information on catch word Gandhi in AIMS software



Search Text (Subject): gandhi

National Archives of India

Total Number of Records Found : 857

Please Check The Record for Requisition

	Sr. No.	Dept	Branch	File Details	Year	Month	Part	Subject
☐	1	Foreign & Political	External	Progs., No. 117, April 1922	1922	4	PART B	Allegation of the ill-treatment of Indian in Mesopotamia.(Published in Mr. M.K. Gandhis paper Young India.
☐	2	Foreign & Political	External	Progs., Nos. 52, March 1921	1921	3	PART B	Information regarding the Gandhi National School at Dehradun & Connection therewith of Thakur Chandar Singh, a Gurkha of Dehra Dun.
☐	3	Foreign & Political	Frontier	Progs., Nos. 193-F, 1931	1931	0		Conversation between Mr. H.W.Emerson home secretary and Mr. Gandhi as bombay on the 6th april 1931.
☐	4	Foreign	Frontier	Progs.,	1931	0		Congratulation of the wali

The AIMS software has integrated system – LAN implementation, six modules and ten sub modules and three levels of security, online processing and retrieval besides report generation and web/internet compatibility. The software was developed with SQL search engine to locate the file details on the basis of catchwords, Records browsing in Departments, Branch, Year, Month and File Details wise, alongwith subject and keywords also tree and flat view retrieval.

There are around 52,00,000 lakhs records for which reference media are to be created. So far reference media has been created for 12,00,000 records and the same is available for retrieval through AIMS software primarily for the benefit of scholars and users of archives in the intranet of the National Archives of India. A database of the entire collection of the Library of National Archives has also been created in the LIBSYS software and is available for retrieval in the intranet of National Archives of India.

Since the Department has a huge collection of microfilms of the document available for the use by scholars, a pilot project of converting the analog microfilm images to digital medium was undertaken in 2000. Around 2,00,000 lakhs images have been created and the AIMS software is being updated in order to have the digital images of the documents on the network, so that frequent use of the document is minimized and multiple viewers could have easy access to the original files in digital form in the intranet in National Archives of India.

It is pertinent to mention that even in countries where e-records or digital documents are created and transferred to their national archival repositories, the records are “future proofed” by transferring the digital image data to microfilm. However, converting digitized data to microfilm as a long-term storage medium involves a notable reduction in quality. Microfilm produced in this way cannot be used for digitization with any guarantee of an acceptable result in comparison to the master negative prepared from original documents. Analog and digital storage forms are thus not fully compatible at present.

Considering the merits and demerits of both microforms and digital media, it is advisable to continue the time-tested process of microfilming for archival records as a means of ultimate preservation. This process is stable, interfaces with digital and analog technologies and is protected against computer viruses. Thus, to preserve archival documents for avoiding the risk of content alteration for indeterminate periods into future, a system that balances the benefits of microfilm media with adequate search facilities and organizational aspects of a digital system with proper indexing for rapid retrieval and online access, appears to be an ideal solution for an archival institution. Till such time, a viable and standardized solution to the digital longevity is developed and accepted as a long term solution of preservation of record data for posterity, we have content ourselves with the traditional medium of preserving in microforms and allied reprographic techniques.

N.S.Mani
Microphotographist
National Archives of India,
Janpath, New Delhi
Email: archives@ren02.nic.in

Electronic Reference Media in National Archives of India – An Overview

The National Archives of India is the custodian of the non-current records of the Government of India and is holding in trust for the use of administrators and scholars. It is an Attached Office of the Ministry of Culture Government of India. It was set up in March 1891 in Calcutta (Kolkata) as the Imperial Record Department and subsequent to the transfer of the National Capital from Calcutta to New Delhi in 1911 it shifted to its present building in 1926. The nomenclature of the Imperial Record Department was changed to National Archives of India after India gained independence on 15th August 1947.

The National Archives of India has in its custody records occupying 30 kilometers of linear shelf space. Most of these records relate to East India Company and the British Raj and are in English. It has also acquired private papers of eminent Indians including Mahatma Gandhi, Rajendra Prasad, Maulana Azad, Sardar Patel, M.R. Jayakar, Field Marshal K.M. Cariappa and others. Approximately 1,73,000 books comprising rare Government reports, proscribed publications, vernacular Native Newspaper Reports, Parliamentary Papers, etc. are as well available in the Library of the National Archives. The Department also has in its custody more than 7000 microfilms of records of Indian interest acquired from other countries. As a long-term preservation measure, National Archives of India has been engaged in security microfilming of important records series and approximately 12,000 microfilms of public records have since been prepared.

The last decade of the 20th century witnessed gradual globalization of archives due to the rapid pace of technological progress and thus necessitated the need to involve the method for facilitating quick retrieval of information. Computerization of archives, which appeared to be the best solution was, then in its infancy, consequently the National Archives of India initiated steps for the development of software for archival use, as there was no internationally recognized and standard software available for archival purposes. At that time most of the countries were laying down specifications for hardware and software according to the nature of their record holdings, which were in their custody. They developed software for Archives in Lotus, dbase-III, FoxBASE, Word Perfect under DOS/Win-3.1/and UNIX Environment as information stored was mostly in textual format (ASCII and flat file).

By 1998, with the further advancement of information technology and availability of new storage medium, the question of creation of data base for reference media of records for quick retrieval of information from the printed reference media was being considered in the National Archives of India as a viable alternative to save time. It was considered economical for research and reference purpose for scholars and users of archives. However many issues were involved and needed to be solved before creation of the database of the reference media of records available in archives. As the storage medium of electronic documents were vulnerable to loss by decay of the medium in which information is stored, they became inaccessible and unreadable especially when the software needed to interpret them or the hardware on which the software runs became obsolete. The advancement in information technology required migration of data to the latest medium frequently. Most digital documents exist only in encoded form, which required specific software to bring their bit streams to action and make them usable. The medium itself disappears from the market; drivers capable of controlling the drives and deciphering the encoding used on the medium, are no longer incorporated in new computers. Upgrading of new computer systems, therefore, requires abandoning the old storage medium, provided an institution still has in its custody documents stored on that medium. The problems of short lifespan of the media and rapid obsolescence have necessitated copying of the digital information into new medium. Though copying is a viable solution, it is likely to corrupt documents due to compression, encryption or ever changing data formats.

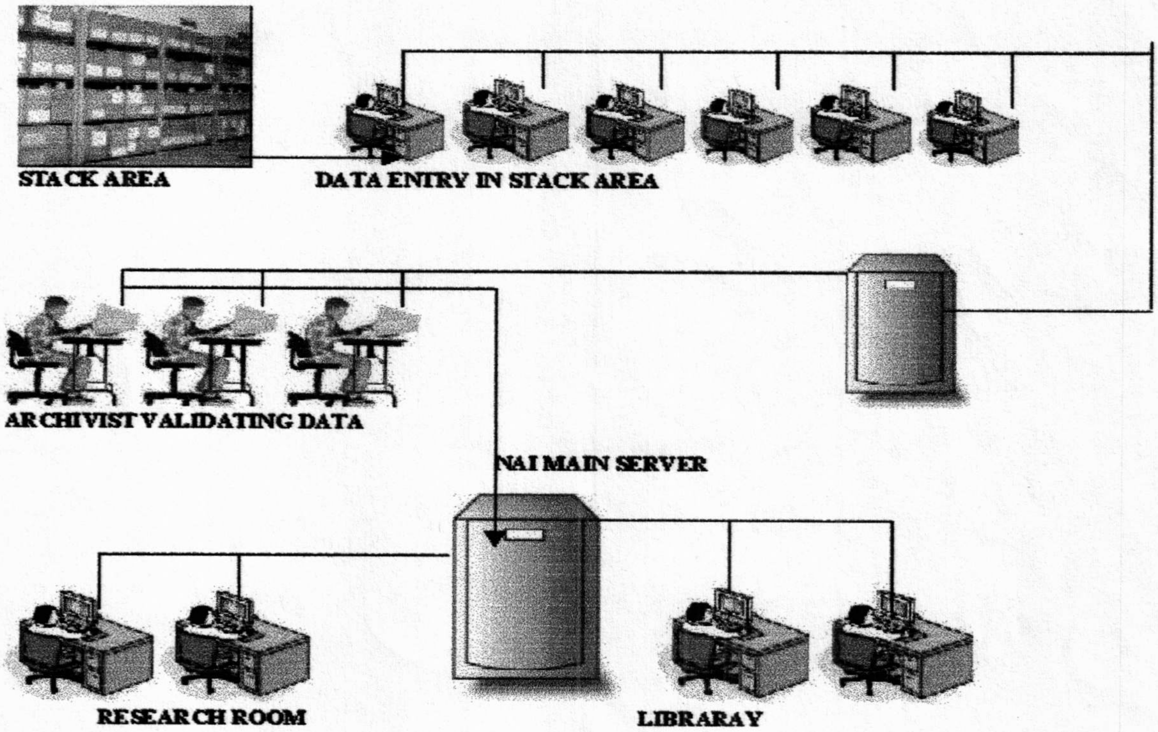
Considering all these aspects, the National Archives of India's first priority was to computerize data of reference media to records for easy retrieval and access for the benefit of scholars and users of archives. Accordingly a tailor made software "AIMS" (Archival Information Management System) was developed by National Archives of India in visual basic front end and MS Office back end. With the introduction of this software, the work of data entry is progressing satisfactorily.

The data entry is now being created from the following formats.

1. Index to press lists, press lists, and transfer lists
2. Printed (type written) calligraphy form (hand written) and manuscripts in various sizes and styles
3. Language - Hindi, English, Persian, Urdu and Sanskrit

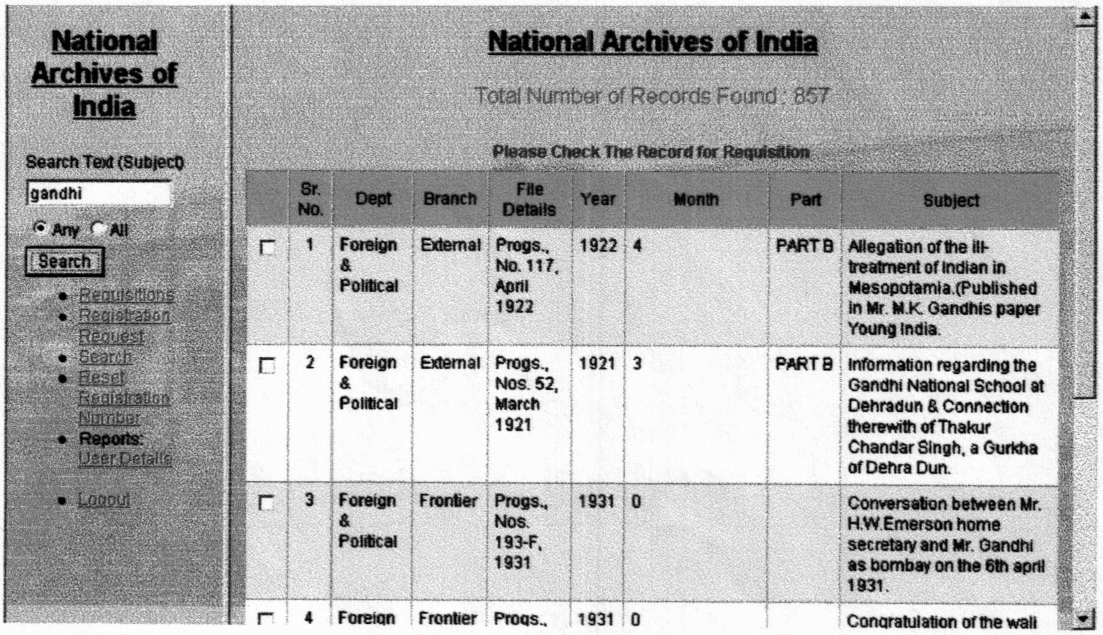
This was, however not an easy task due to a number of problems faced, viz., diverse forms and formats of records; no fix data structure; cross linking of records; no existing standards for computerization; multiple keywords within the subject field.

Graphic Representation of Data Entry and Data Validation for Reference Media of Records in National Archives of India, New Delhi.



Simultaneously, the National Archives of India designed an integrated system solution for the Archival Information and Management System (AIMS) software on intranet ware. The software was developed with web browser based access and having facility to upgrade for web/intranet and WAN deployment. The software had flexible modular design and a package was developed to facilitate computerization of indexes to records. Besides, AIMS software was customized to access the information related to an important historical event, chronological span, and activity of a personality, political leaders and much more.

Retrieval of information on catch word Gandhi in AIMS software



Search Text (Subject): gandhi

● Any ○ All

National Archives of India

Total Number of Records Found : 857

Please Check The Record for Requisition

	Sr. No.	Dept	Branch	File Details	Year	Month	Part	Subject
☐	1	Foreign & Political	External	Progs., No. 117, April 1922	1922	4	PART B	Allegation of the ill-treatment of Indian in Mesopotamia.(Published in Mr. M.K. Gandhis paper Young India.
☐	2	Foreign & Political	External	Progs., Nos. 52, March 1921	1921	3	PART B	Information regarding the Gandhi National School at Dehradun & Connection therewith of Thakur Chandar Singh, a Gurkha of Dehra Dun.
☐	3	Foreign & Political	Frontier	Progs., Nos. 193-F, 1931	1931	0		Conversation between Mr. H.W.Emerson home secretary and Mr. Gandhi as bombay on the 6th april 1931.
☐	4	Foreign	Frontier	Progs.,	1931	0		Congratulation of the wali

The AIMS software has integrated system – LAN implementation, six modules and ten sub modules and three levels of security, online processing and retrieval besides report generation and web/internet compatibility. The software was developed with SQL search engine to locate the file details on the basis of catchwords, Records browsing in Departments, Branch, Year, Month and File Details wise, alongwith subject and keywords also tree and flat view retrieval.

There are around 52,00,000 lakhs records for which reference media are to be created. So far reference media has been created for 12,00,000 records and the same is available for retrieval through AIMS software primarily for the benefit of scholars and users of archives in the intranet of the National Archives of India. A database of the entire collection of the Library of National Archives has also been created in the LIBSYS software and is available for retrieval in the intranet of National Archives of India.

Since the Department has a huge collection of microfilms of the document available for the use by scholars, a pilot project of converting the analog microfilm images to digital medium was undertaken in 2000. Around 2,00,000 lakhs images have been created and the AIMS software is being updated in order to have the digital images of the documents on the network, so that frequent use of the document is minimized and multiple viewers could have easy access to the original files in digital form in the intranet in National Archives of India.

It is pertinent to mention that even in countries where e-records or digital documents are created and transferred to their national archival repositories, the records are "future proofed" by transferring the digital image data to microfilm. However, converting digitized data to microfilm as a long-term storage medium involves a notable reduction in quality. Microfilm produced in this way cannot be used for digitization with any guarantee of an acceptable result in comparison to the master negative prepared from original documents. Analog and digital storage forms are thus not fully compatible at present.

Considering the merits and demerits of both microforms and digital media, it is advisable to continue the time-tested process of microfilming for archival records as a means of ultimate preservation. This process is stable, interfaces with digital and analog technologies and is protected against computer viruses. Thus, to preserve archival documents for avoiding the risk of content alteration for indeterminate periods into future, a system that balances the benefits of microfilm media with adequate search facilities and organizational aspects of a digital system with proper indexing for rapid retrieval and online access, appears to be an ideal solution for an archival institution. Till such time, a viable and standardized solution to the digital longevity is developed and accepted as a long term solution of preservation of record data for posterity, we have content ourselves with the traditional medium of preserving in microforms and allied reprographic techniques.

N.S.Mani
Microphotographist
National Archives of India,
Janpath, New Delhi
Email: archives@ren02.nic.in

राष्ट्रीय अभिलेखागार
NATIONAL ARCHIVES OF INDIA

सत्यमेव जयते

अपनी विरासत का संरक्षण
Preserving our Heritage

नेशनल कमेटी ऑफ आर्किविस्ट्स की बैठकों में प्रस्तुत आलेख
Papers presented at meetings of National Committee of Archivists
(1954 - 2007)

नई दिल्ली
NEW DELHI
2008

10. Care and Conservation of Newspapers by Dr. Yashodhara Joshi.............. 61

11. Pest Management in Archives and Libraries by Dr. Yashodhara Joshi......... 67

12 Digital Archives by Ajit Kr. Dey............... 74

13. Recent Trends in Digitalization of records by N.S.Mani..................... 79

14. Essentials for effective conservation of documentary heritage in 21st century by Dr.C.L.Prajapati.................... 84

15. Development of Indigenous Repair Techniques: Indigenous substitute for solvent lamination by Dr. Yashodhara Joshi....................... 89

16. Efficacy and lifetime of Polymer U.V. Films in protecting deterioration of documentary Heritage against U.V. Radiations by Dr. C.L.Prajapati................. 96

17. Preservation of Cultural Heritage in Karnataka State Archives through Digitisation by Dr.Usha Suresh........... 100

18. Electronic Records in Archives: A challenge for preservation by N.S.Mani.................... 106

Appendices

		Pages
Appendix 1.	Draft scheme on "Manual of Archives' by **Dr. V.K.Bawa...**	114
Appendix II.	Note on minimum requirements for setting up a Conservation Research Laboratory	117
Appendix III	Note on minimum requirements of a State Record Office......	128
Appendix IV.	Constitution of National Committee of Archivists......	136
Appendix V.	List of Meetings of National Committee of Archivists (1954-2007).......	138

ELECTRONIC RECORDS IN ARCHIVES – A CHALLENGE FOR PRESERVATION*.

by
N.S. Mani, Microphotographist,
National Archives of India.

The rapid technological advance is gearing towards a spectacular change in accumulation of Archives in the years to come. The business of government is facilitated by the ever-growing application of electronic tools for communicating, processing transactions and managing information. Digital documents could be seen in every sphere of government, legislative, executive, administrative, regulatory and judicial agencies. As a result Archives are witnessing the birth of an electronic government. These electronic or digitized records should be preserved and made available in readable format, whenever needed.

Towards 21st Century, many digital documents would be created digitally as email, word processing or spreadsheet files, sometimes routed as email attachments as also in online forms. Some of the old and existing documents are converted from paper documents to digital formats by scanners and other devices.

The aim of Archives is towards long-term preservation of record material of the past and present for the benefit of distant future. The Archives, which have a mammoth task in safeguarding the documents/records, are now confronted with the task of long time preservation of electronic records. The need of the hour is to evolve a long-term preservation policy for preserving traditional archives. The task involves finding suitable, reliable and stable structures for carrying preservation activity, which ensures that data is properly selected, stored, with its logical and physical integrity being maintained over the period of time so that they could be accessed and disseminated in time with the technological changes with out jeopardizing security and authenticity. Archivists will have to take action to ensure that information available with them is rendered usable in the future.

***Presented at the 43rd Meeting, ,Shimla 24– 25, September,2007**

Archives, therefore, should be guided by long-term management and preservation aspects and strive to ensure that the preservation methods and techniques are upgraded periodically. The digital file formats in which the electronic records are stored in a computer include many commercially based formats such as Word, Excel, Access and Power Point formats.

The details of current commercial formats are proprietary and difficult and sometimes impossible to obtain when the software is upgraded or changed in the fast ever growing information technology. Consequently, the format is more likely to be unavailable when we would actually access it, since that format no longer remains the "current" version of the developer/proprietor.

Conventional preservation policy in Archives are primarily confined to the management and upkeep of collections of record, i.e. paper, books, photographs, etc., and dissemination of information to scholars and users of Archives through traditional means, i.e. xerox copies, publications, microfilms, microfiche, etc. The increasing use of digital technology for creating and communicating information in Government Departments since the late 1980's and their transfer to Archives after 25 years as per the Public Records Act and Rules, in digital media, substantially requires different conservation methods and a well knit preservation policy. The current policy does not lay stress on immediate and continued access to important digital data. Rapidly changing institutional preservation policy is the need of the hour for the survival of archival data for future generations.

Digital preservation comprises different processes to ensure that digital resources remain accessible, usable and understandable in future. Long term preserving digital resources does not guarantee a responsible development, which could cope with constant changes brought about by the information boom. It is, therefore, a major challenge to organize a correspondingly broad-based and diverse discussion incorporating views from all concerned especially technology experts in information science towards the development of a definitive preservation policy for Archives.

Archives in the digital age, must meet the requirements of information boom in new digital media, in preservation policy to fulfill their mandate as guardians of the documentary heritage. Archivists would have to identify the key problems for continued access to digital collections, define critical issues and recommend actions to eliminate barriers to instant solutions besides consider other technological answers to digital preservation.

Digitization

Digitization is the translation of analog text, sound or images into digital format and results in the creation of digital surrogates containing similar information as the original. This differs from objects born in digital format. Digital collections are in a format which could be delivered electronically anywhere and anytime, they can be viewed at the size and resolution necessary for the particular task, they can be downloaded and saved for later use and have the potential to be viewed or used simultaneously,

Text or images, typically printed, manuscript or graphic works on traditional substrates such as paper or parchment, can be digitized by the use of optical scanners or digital photography. In the case of optical scanners, the document is placed in or on the scanner and the scanner translates the image into digital code. Scanners could be used to scan text in two different ways: the text can be scanned as an image; or it can be scanned using optical character recognition (OCR) software to Save it as a text file. OCR software digitizes text as alphanumeric characters, which can be further edited by word processing software programmes. The second option scans the text and saves it as an image where characters are treated as areas of light, shade or colour. Manipulation of the resulting image file is performed only through graphics or illustration software (such as "Adobe Photoshop") and text editing is not possible. Digitization by camera similarly captures information as an image, not as text.

1. Born Digital

For a record to be "born" digital format, it is to be created by a computer system. Digital technology has brought forth many significant advantages to users in many fields. The interest of creator of the digital documents in the computer is not necessarily long-term preservation or access to the information. Archives see quite a different set of advantages and disadvantages associated with digital technologies than do the creator. The disadvantages of digital media and technology, and the challenges they present to heritage institutions, are related to the long-tem preservation.

Another important characteristic of digital electronic records is their logical format. Bits and bytes cannot represent real-world things without application of additional means through which this code could be rendered accessible and understandable as informational content. Computer software programmes provide the means to organize and package data into logical formats programmed to represent information in specific ways. Logical formats are usually proprietary to their creators (software companies) and more often not completely compatible with software programmes other than that in which they were originally created.

2. Electronic Textual records

Electronic textual records include all types of documents created by computer systems (typically word processing software) that would have previously created through various printing or manuscript techniques (e.g., typewriter, adding machines, letterpress, hand writing, etc.). Digital textual records are produced by computer authoring systems in a variety of logical formats (usually proprietory), and on various media (e.g., magnetic tape or disk, optical media, computer memory chips). Common proprietory formats include, Corel, Word Perfect or Microsoft Word. Non-proprietory or "open" file formats include, the ASCII format, the Adobe" PDF format, and markup language formats used for exchange over the Internet (e.g. HTML, XML).

3. Databases and Spreadsheets

These records are created as tables or indexed collections of data (numeric or alpha-numeric) image by software programmes. As with textual records they typically exist in collections as digital files in proprietory or non-proprietory formats on magnetic or optical medium.

The adoption of digitization as a preservation copying strategy is also pitted against formidable and well established traditions. The standard format preservation for microfilming and photography has set this medium with long-term potential. Recent technical expert still support microfilming as the preservation reformatting option of choice for analog originals.

4. Migration

Migration ensures the periodic transfer of digital information from one software/ hardware configuration to another available medium in vogue, or from one generation of computer technology to a subsequent generation of information technology. If there is a potential for loss of valuable information during migration then institutions would have to determine what would be a total acceptable quantum of loss (risk management) with the primary goal of access in mind, which should be to kept as watchword or an evolution of a definite policy to be evolved.

The various aspects of digital preservation, which have to be borne in mind and relate to the following: -

1. How is preservation of digital object different than preserving an analog object?
2. How does preservation depend on access?
3. What are the essential attributes of a digital object that should be preserved?
4. How should you manage preservation of the enormous volume of digital objects created by governments and transferred in disk/CD formats to archives for preservation?

The professionals in archives are often faced with complex and numerous related issues confronting their established tradition, practice and ethics. Digital preservation represents a significant new paradigm for archivists for archival collections and their management.

In view of the electronic records being generated with speed, there appears to be a need for evolving preservation policy at the national level as first priority. To meet this challenge Archives should evolve and develop policy guidelines, for the new techniques/ procedure.

Objectives

The objectives of a digital preservation policy, along with its purpose and scope, are consistent with those objectives typically identified for traditional media but are broadened or extended to account for a new medium and process, and work out the priorities for their use and access. Digital preservation typically centers on the choice of interim storage media, the life expectancy of a digital imaging system, and the expectation to migrate the digital files to future systems while maintaining full functionality and the integrity of the original digital system

Archivists' concern the world over, relate to the authenticity and integrity of data are implied and therefore, deserve special attention. In archives, maintaining the value and importance of digital holdings ensures that the information accessed by the user is the same information that was initially acquired and described and which still relays its context and structure. As stated earlier, the care of manipulation of digital data, either intentional or accidental, leads to deliberate alteration or deletion of information without any trace or evidence.

The principal objective for archives should therefore be the document management of the digital record over a period of time in order to guarantee the reliability of its content and meaning. An objective of digital preservation policy should be to clarify and extend the application of ideas to new medium or to introduce new concepts keeping in view the necessary adherence to the history, provenance of archives, etc.

Digital preservation policy should include and highlight the timely provision of advice or provide support to the areas of collections, acquisition and operations. This is crucial in the digital world as technological obsolescence imparts urgency to the areas, acquisition and intellectual control activity of the concerned archival institutions. The institution should bring the object under "preservation control" (ensuring current and future access) with the technology necessary to access its informational content. This issue becomes critical when the institution has a considerable backlog of programmes relating to acquisitions or preservation due to resource crunch or other delaying factors.

Traditionally a successful track record is necessary to launch any new endeavour. In the case of digital preservation they provide an effective launch pad for the development of a new policy model for the information age, which builds on existing knowledge and incorporates added layers of technological and professional concerns. Thus, traditional preservation policy offers such a foundation as it represents a mature approach to the preservation of varied collections of cultural heritage and reflects advances in professional ethics and institutional vision.

Established preservation principles relating to respect for the integrity of the original, minimal intervention, reversibility, as well as concepts of preventive preservation, custody and access can only partially support the demands of digital collections.

Traditional preservation principles are so substantially transformed when applied to digital materials that in practice they eventually fail to describe the real existing scenario in a meaningful way. Frequent debates on the nature and relevance of preservation in a digital world especially the role of archivists, librarians, curators, managers, politicians, consumers, manufacturers, is necessary to clarify principles which encompass the preservation of all cultural resources, digital or other formats.

While formulating preservation policy for Archives the following must be taken care in to account: -

1. Co-operation–national level management of Digital Archives/Repositories – the State Archives have a major role to play in it.
2. Redundancy of digital equipments hardware as well as software
3. International standard interoperatability

The policy should spell out answers and solutions for digital media longevity, rates of deterioration of digital objects, storage conditions and disaster recovery. The retention period of the paper records and electronic records should also be looked into so as to ensure an appropriate retention period of the electronic records is fixed. This could be deliberated in high-level techno professionals' forum representing the archival fraternity alongwith information technological experts.

The subject of digital preservation is still in the primary domain of the information technology professionals. The acquisition, preservation and management of objects in new medium pose a challenge to Archives. Besides there has been an increase in the quantum of information being created digitally in Government. The speed with which new technologies to preserve information are applied have not kept pace with its creation.

It was reported in 2003 that the number of internet users worldwide has been increasing 200 % every year. The rapid pace of technological development and obsolescence, threatens vast amounts of information of being "stranded" on tradition or old medium as new and better techniques supersede them. Current approaches to the preservation of important digital information rely on techniques, which must be applied within the useful life span of medium and their dependent technologies (software, hardware). Apathy to look for trusted means and methods of digital preservation would certainly exact a stiff, long-term cultural penalty.

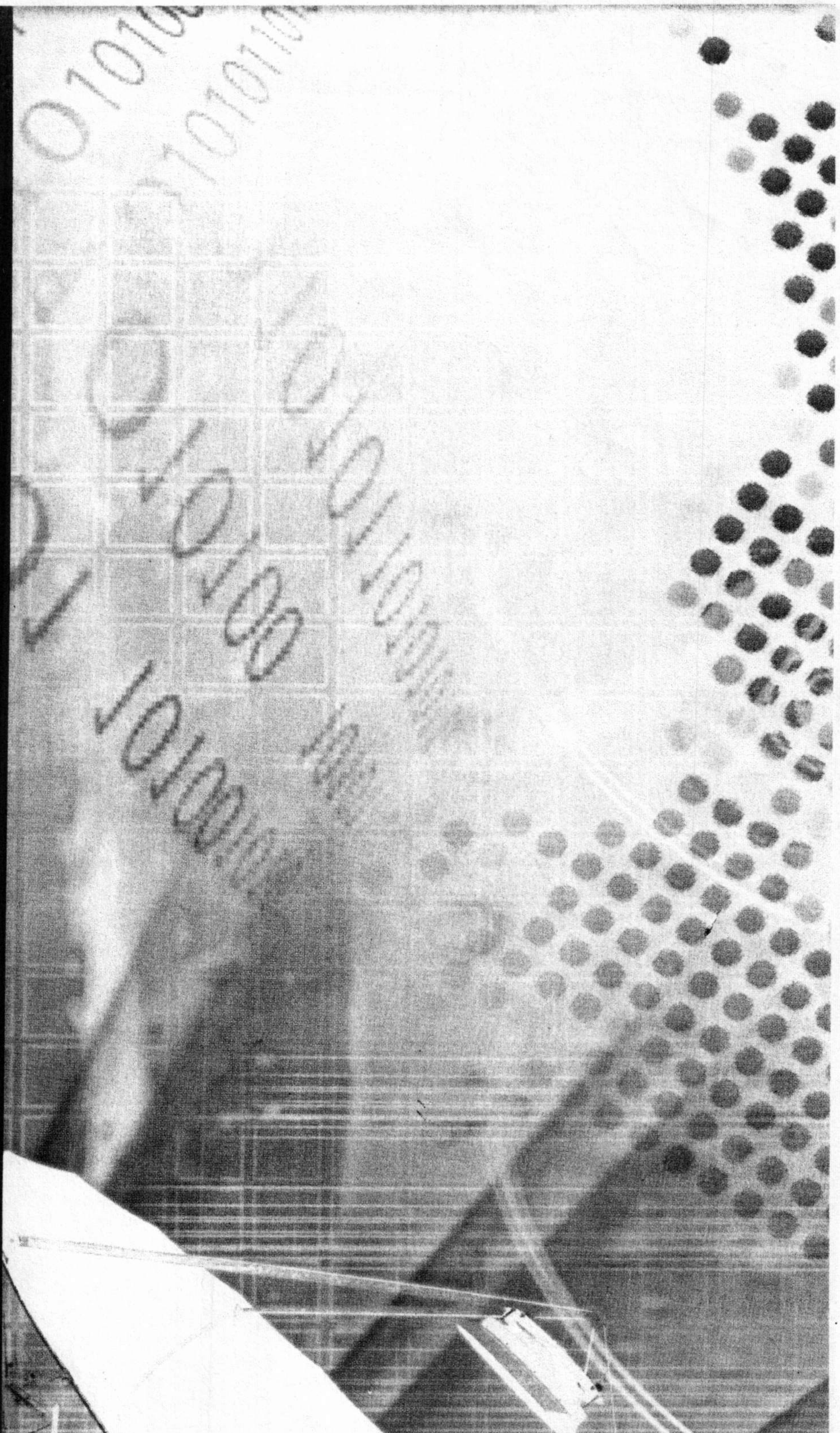
Adapting preservation policy to the digital age is dependent on the coordination of a complex set of layered determinants of a cultural, a technological and a professional nature. This requires an appreciation of the larger cultural context for framing long term institutional policy and a significant knowledge of digital technology systems and trends. This means evaluating current preservation practices and principles against a new digital paradigm, and finally it demands new attitudes to the preservation of cultural heritage collections in digital form. However the financial issues can be deliberated upon once a consensus is evolved on the policy for digital long time preservation. Financial issues are important and are likely to be supported by information experts' advice on cost-effectiveness of hardware and software. This measure would ensure smooth migration on the available latest technology during every ten years in a planned and phased manner.

In the end I would like to emphasise that digital preservation is the means of access to collections and an instrument for the realization of national cultural policy goals.

ISSN No. : 0974-0643
Vol. 29 January 2009 No. 1

DJLIT

DESIDOC Journal of Library & Information Technology



Special Issue on
E-Learning

DEFENCE SCIENTIFIC
INFORMATION &
DOCUMENTATION CENTRE

Defence Research &
Development Organisation
Ministry of Defence
Metcalfe House, Delhi-110054

DESIDOC Journal of Library & Information Technology, Vol. 29, No. 1, January 2009, pp. 69-71
©2009, DESIDOC

Special Paper

Digitisation: Preservation and Challenges

N.S. Mani

National Archives of India, Janpath, New Delhi-110 001
E-mail: archives@nic.in; nsmani53@hotmail.com

ABSTRACT

Digitisation involves conversion from analog to digital. With the boom in the information technologies and also with the changes in the retrieval of information need arises to convert the heritage into digital form. This paper discusses Digitisation, its need and how to preserve the digitised documents.

Keywords: Digitisation, preservation, conversion, resolution

1. INTRODUCTION

Digitisation is the process of converting an analog signal into a digital signal. Digitisation converts an image into a series of picture elements or pixels (little squares that are either black or white), a specific shade of gray (grayscale) or colour. Each pixel is represented by a single or series of binary digits, either 1s or 0s. Many computer programs like drawing, illustration, 3D model, computer-aided design/computer-aided manufacturing, and architectural design use vectors—arrows of direction, points, and lines that define shapes, as compared to the individual picture elements used to represent a raster image. One important part of digitising is the type of imaging sensor used. These image sensors/detectors can be compared to the film used in photography. The common digital imaging sensors are:

(i) Charged Coupled Devices, or CCDs, used in flatbed scanners and digital cameras.

(ii) Photomultiplier tubes, or PMTs, used in drum scanners.

(iii) Complementary metal oxide semiconductors, or CMOS chips, used in low-end flatbed scanners and low-end digital cameras.

The most common sensor used in all types of scanners and digital cameras is CCD. The photo scanners and digital cameras is CCD. The photo multiplier tube, or PMT, is used only in drum scanners for the graphic arts or prepress market, i.e., printing and publishing. CMOS chips do not produce the same image quality as compared to CCDs and cannot match the resolution of CCDs. Now-a-days, CMOS chips are used only in low-end digital cameras and scanners. In scanners and digital cameras, a lens or set of lenses are used to focus an image onto the sensor—a CCD, PMT, or CMOS chip. Sometimes people refer to digital imaging as lens-less. This is not true. Without a lens to focus the light, digital images would be blurry, just as in photography. All scanners have lenses. In most cases, the design of the scanner hides the lens (es) within the body of the scanner. Scanners also have built-in light sources to illuminate the items being scanned. The light is either reflected (as with documents or photographic prints) or transmitted (as with microfilm, photographic negatives, or colour transparencies) by the item being scanned, and the image is focused by the lens (es) onto the imaging sensor. In the case of CCDs, the light falls onto the little light sensors or photo diodes on the CCD. These sites generate an electrical current, or voltage.

2. ANALOG-TO-DIGITAL CONVERSION

Once a digital image has been created and stored in any media, there is a corresponding digital to analog conversion that allows the computer to present the image in a human readable form on either a display or printer. Displaying an image on a computer monitor or printing the image on a printer needs analog representation of a digital image.

2.1 Basic Image Measures

There are three important measures of every static digital image:

The paper was presented in the National Conference on Digitisation and Preservation organised by DESIDOC at Metcalfe House, Delhi-110 054.

Resolution: The number of dots, or pixels (picture element), used to represent an image. This is always given as a measure of linear or area density (e.g., 300 dots/inch). Resolution, or spatial frequency, is the number of times an image is sampled during the scanning process. Resolution—the number of pixels in an image can be described in a number of ways like dots per inch (dpi) or pixels per inch (ppi).

Pixel Bit Depth: This measure defines the number of shades that can actually be represented by the amount of information saved for each pixel. These can range from 1 bit/pixel for binary (fax type) images to 24 bits per pixel in high quality colour images.

Colour: There are many ways to represent, compress, and distribute color images. Suffice it to say that the smaller the image files size, the less accurately it renders the original image.

The key to a successful digital conversion program is the relationships among three concepts:

(i) The purpose digitisation is going to serve.

(ii) Original document's condition and status.

(iii) Technology tolerance for the analog-to-digital conversion.

Information in digital form is more fragile. The recent survey of the Web shows that the overall proportion of Internet content is commercial and only six percent is educational or informational. Digital images have become commonplace in libraries and archives. With serious effort to ensure long-term access to today's digital image files, organisations like library, archives, historical society, or museum are earmarking a certain percentage of their budget for digital preservation without which the risk of loss is tremendous.

3. PRESERVATION

Digital preservation refers to the various methods of keeping digital material alive for the future. Digital preservation typically centres on the choice of interim storage media, the life expectancy of a digital imaging system, and the expectation to migrate the digital files to future systems while maintaining both. Digital imaging technologies create an entirely new form of information from traditional documents. It is not simply another reformatting option in the preservation tool kit. Digital imaging involves transforming the very concept of format, not simply creating a faithful reproduction of a book, document, photograph, or map on a different medium. The power of digital enhancement, the possibilities for structured indexes, and the mathematics of compression and communication together fundamentally alter the concept of preservation in the digital world. These transformations, along with the new possibilities they place on information professionals, force us to transform library and archival services and programs in turn.

Preservation is not just only for paper/palm leaves. In the digital imaging environment it is indeed a matter of challenge to define what preservation means? The digital world poses significant challenges to, but does not eliminate the need for responsible, effective preservation activity. Digitisation offers many advantages to organisations with large collection of documents/maps/microforms/etc., Information from these can be accessed by users directly without the human intervention. Information content in digital form can be retrieved by users remotely.

Digital image quality is improving day by day with more advancement in research on this field. The brittle, old, and fragile originals need not be consulted again and again for research and the digital form of these can be accessed by many at a single point of time for research. The power of full-text searching and sophisticated, cross-collection indexing affords users the opportunity to make new uses of digitised resources.

Digitisation of documents/books or cultural resources requires a deep and longstanding Organisational commitment in terms of investment in equipment and software as compared to the traditional preservation. Wherever large scale Digitisation programme are being undertaken it should be a prerequisite to decide definitions and standards for digital preservation and policy for migration.

In the digital world, preservation is the creation of digital products worth maintaining over time. The persistence of digital products requires careful attention to the maintenance of content (the bits and bytes) functionality (how the bits work in a system). Preservation in the digital world is not absolute, but depends on the continuing transformative impact of the digital product on the information work of end-users. Acquiring an imaging system to enhance access to library and archives materials is as simple as choosing the combination of off-the-shelf scanners, computers, and monitors that meets immediate functional specifications.

With time, most of these small-scale, pilot projects will fade away quietly and the initial investment will be lost. The costs of maintaining these systems become apparent, as vendors go out of business, and as patrons become more accustomed to remote-access image databases and the latest developments in the field. This commitment is a continuing one, decisions about preservation cannot be deferred in the hope that technological solutions will emerge like a medieval knight in shining armor. An appraisal of the present value of a book, a manuscript collection, or a series of photographs in its original format is the necessary point of departure for making a judgment about preservation of the digital image version.

Before initiating a digitisation programme, the following should be kept in mind

- The condition and status of the document being converted.
- The capabilities of the present technology for digital conversion, and
- The method by which the digital images will be utilised.

The common application of Digitisation in an archive or library is digital copies that can be used for ready reference in lieu of consulting the original documents. For Digitisation the use of an intermediate film or paper copy to facilitate the scanning process may or may not be necessary or advisable. Ultimately, the purpose of digital image products is determined by the uses to which they will be put, while preservation of original source documents must be determined by their specific preservation needs.

A major challenge in choosing methods from analog to digital conversion is obtaining an in-depth understanding of the source material and the techniques being used for the same. The following should be taken in to account before undertaking the job:

- Format of the source (including size of object, its structure, and its physical condition).
- Physical condition and its impact on the ability of the item to be handled during the conversion process.
- Level of detail (including the colour of the paper ink used etc.).

All conversion tools have limitations in terms of the size of source documents they can handle with a given level of digital resolution. "You get what you pay for" typically applies in the selection of scanning equipment. The principles of preservation of digital documents are the same as those of the analog world, i.e., longevity, choice, quality, integrity, and accessibility. When we talk of preservation, archivists not only insist for alkaline paper standards, archival quality microfilm, mass deacidification, and more rugged magnetic media, but also issues surrounding the life expectancy of storage media, the very concept of permanence. The commitment to quality standards, do it once, do it right, permeates all preservation activity, including library binding standards, archival microfilm creation guidelines, conservation treatment procedures, the choice of supplies and materials, and a low tolerance for error. The evolution of preservation microfilming as a central strategy for the bulk of brittle library materials has placed the quality of the medium and the quality of the visual image on an equal plane. In the pursuit of quality microfilm, compromise on visual truth and archival stability is dictated largely by the characteristics of the item/material chosen for the preservation.

The control of digital quality standards is possible now, just as it is for microfilm. In the digital world, preservation is the action and access is the thing—the act of preserving access. A more accurate construct simply states "preserve accessibility". Librarians and archivists can influence vendors and manufacturers to provide new equipment that is backward compatible with existing systems. This capability assists image file system migration in the same way that today's word processing software allows access to documents created with earlier versions. Procedures manuals (or instruction sheets) for selection, handling, scanning, metadata creation, and quality control flowcharts or other workflow diagrams, data element list the contents.

4. CONCLUSION

Since libraries and archives are actively integrating digital technologies into acquisition, cataloging, systems, and even preservation departments, perhaps it is legitimate to consider these activities as costs that lie outside of a conversion project. Copyright does not apply to all works, and it does not last forever. It is intended as a limited monopoly permitting authors to profit from their creative efforts and, eventually, for the public to be able to freely use creative works with the idea that such fluid use will inspire new creation benefiting society at large. When a creative work is not subject to copyright protection, it is referred to as being in the public domain. If a use is a fair use, then one may copy an otherwise protected work without permission from the copyright owner and without the use being deemed an infringement. Other countries have a similar concept often called "fair dealing," but this is generally more restrictive and laden with administrative requirements.

In view of the multifocussed challenges being posed by Digitisation, the Government of India has very recently formed a forum–National Digitisation Preservation Programme [NDPP] to formulate different programmes and polices for the same.

Indo-US Science and Technology Forum

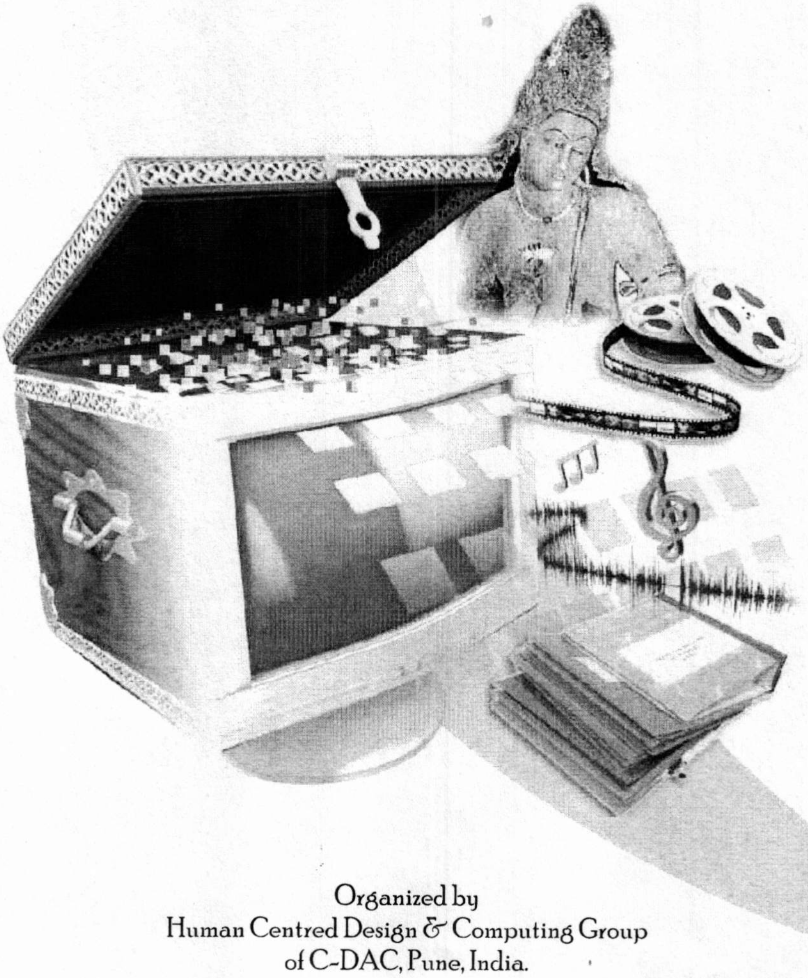
सी-डैक
CDAC

Proceedings

Indo-US Workshop on International Trends in Digital Preservation

March 24-25, 2009

This initiative is a part of the National Digital Preservation Programme
of Department of Information Technology, Government of India.



Organized by
Human Centred Design & Computing Group
of C-DAC, Pune, India.

www.cdac.in

ARCHIVES AND DIGITAL PRESERVATION: VARIANT CONTOURS

N.S. Mani
Janpath, New Delhi –110001
Telephone number, incl. country code
nsmani53@hotmail.com

ABSTRACT

Digital preservation of archival records in India has engaged the attention of policy planners in governmental administration. A beginning has been made in this direction during the 20th century. During the 21st century, the problems of digitization of descriptive standards along with preservation of digitized data are going to assume gargantuan proportions.

This would lead to laying greater emphasis on records creation on digital format. The creators as record managers and archivists as a custodian of this novel format would have to devise strategies to evolve viable logistics for a long term planning along with intellectual exploitation of contents on digital format.

Besides, in order to control the implosion of this format variety of inputs to control their explosion and exploitation would have to be devised for ensuring authentic digital sources before they are preserved for posterity.

This paper would be a stimulant for further research on the need of digital preservation in all key sectors of the governmental administration, for harnessing this novel source by striking a perfect balance between the creating agencies, custodians and users of records in the backdrop of ever-growing concern for preserving national heritage of India. The Indian experience as gathered from National Archives of India in the field would open up new vistas of insights in future for the ones who are waiting in horizon to join the digital world of archives.

1. INTRODUCTION

The aim of Archives is towards long-term preservation of record material of the past and present for the benefit of users in the distant future. The Archives, have a mammoth task in safeguarding the documents/records, are now confronted with the task of long time preservation of electronic records. The need of the hour is to evolve a long-term preservation policy for preserving traditional archives. The task involves finding suitable, reliable and stable infrastructures for carrying preservation activity, which ensures that data is properly selected, stored, with its logical and physical integrity being maintained over the period of time for being accessed and disseminated perfectly in tune with the technological changes without jeopardizing security and authenticity. Archivists will have to take timely action in tandem to ensure that information available with them is rendered usable in the future.

The rapid technological advances are gearing towards a spectacular change in accumulation of Archives in the times to come. The business of government is facilitated by the ever-growing application of electronic tools for communicating, processing transactions and managing information. Digital documents could be seen in every sphere of government, legislative, executive, administrative, regulatory and judicial agencies. As a result Archives the world over is a witness to the birth of an electronic government. These electronic or digitized records of this generation should be preserved and made available in readable format, whenever needed.

Towards 21st century, many digital documents would be created digitally as email, word processing or spreadsheet files, sometimes routed as email attachments and also in online formats. Some of the old and existing documents are converted from paper documents to digital formats by scanners and other devices.

Archives, therefore, should be guided by long-term management and preservation aspects and ensure that the preservation methods and techniques are upgraded periodically. The digital file formats in which the electronic records are stored in a computer include many commercially based formats such as Word, Excel, Access and Power Point formats. The details of current commercial formats are proprietary and difficult and sometimes, impossible to acquire especially when the software is upgraded or changed in the backdrop of changing information technology. Consequently, the format is more likely to be unavailable when we would want to actually access it, since that format no longer remains the "current" version of the developer/proprietor.

The ever increasing use of digital technology for creating and communicating information in Government Departments since the late 1980's and their transfer to Archives after 25 years as per the Public Records Act and Rules of Government of India, in digital media, requires different conservation methods and a well knit preservation policy. The current conventional policy of management and upkeep of archival collections through xerox copies, publications, microfilms and microfiche does not lay stress on immediate and continued access to important digital data. Rapidly changing institutional preservation policy is the need of the hour for ensuring survival of archival data for future generations.

Digital preservation comprises varied processes to ensure that digital resources remain accessible, usable and understandable in future, which has the capacity to cope with constant changes brought about by the information boom. Therefore, a major challenge to initiate a correspondingly broad-based and diverse discussion incorporating views from all concerned especially technology experts in information science towards development of a definitive preservation policy for Archives. Archivists would have to identify the key problems for continued access to digital collections, define core issues and recommend action to eliminate barriers to instant solutions besides consider other technological solutions to digital preservation.

2. NATIONAL ARCHIVES OF INDIA

The National Archives of India is the repository of the non-current records of the Government of India which are being held in trust for the use of administrators as well as scholars. It is an Attached Office of the Ministry of Culture and is an implementing nodal agency of the Government of India for the Public Records Act, 1993 and Rules

The total holding of National Archives comprises the following:

1. 38,75,332 files
2. 64,221 volumes
3. 1,10,332 maps and cartographic items
4. 3,601 bills assented to by the President(s) of India
5. 1,065 treaties
6. 2,442 rare manuscripts
7. Microfilms: 35,000 Negative and Positive

approximately. [contains 4.9 million pages approximately]

8. Microfiches: 9000 [0.9 million pages]
9. Digital Images: 2, 60,000 [TIFF and PDF format

3. SECURITY MICROFILM PROGRAMME

For preserving the cultural heritage of the country National Archives of India has taken up gigantic task of creating a backup of the originals by microfilming the entire collection of documents and keeping them away from the originals in its Regional Office at Bhopal wherein the temperature and humidity is controlled 24 x 7, 365 days. This practice is being followed internationally, as the life expectancy of film medium is more than 500 years in comparison to the other media.

Libsys software in Linux platform is being used for retrieval of information from approximately 1.5 lakh books / periodicals / journals in National Archives of India. The system provides Online Public Access of catalogues of National Archives of India Library.

In July 2000 National Archives of India had initiated a pilot project for digitization of rare manuscripts with the help of Sony Digital Camera. This Unit was established in collaboration with Department of Science and Technology and National Institute of Advanced Studies, Bangalore. Rare manuscripts, viz. Bhagwadgita, Ramayana, Mahabharata, etc. have been digitized and stored in CD medium.

Besides in 1998, National Archives initiated and is continuing the digitization of converting analog microfilms into digital format by using the Microfilm Scanner. The main idea behind this conversion has been to facilitate the access of the fragile documents which have been microfilmed for scholars and users consulting Archives online without the handling the original documents. The microfilm images are being converted into PDF format

It may be worthwhile mentioning that National Archives of India has already taken an initiative to preserve the contents of the current important files which are not covered under the Public Records Act, 1993. This relates to the microfilming of `A` category records.

In order to prevent the content loss of the current files which are very important, the National Archives of India has commenced microfilming of all important `A' category files in the storage area of the records in all Ministries/ Departments/ Offices of the Government of India. The idea behind this task is to preserve the negative or master copy of the microfilm in the National Archives of India and preserve the positive microfilm of `A` category files in the concerned Ministry/ Department/Offices.

Online Exhibitions on Internet:

The following exhibitions organized by National Archives of India are now online in the National Archives website. [www.nationalarchives.nic.in]

1. Commemorating 150 years of Presidency Universities [1857-2007]
2. Hindi Desh Se Pardes Tak

4. STANDARDS AND MEDIUM USED IN NATIONAL ARCHIVES FOR DIGITIZATION

Formats	**Scanning**	**Stores in**
Microfilms and Microfiche	300dpi	HDD, Removable HDD, DVD, CDs
Photographs	300dpi	DVDs [Duplicates stored in different locations.]
Books and Printed materials	300dpi	DVDs

4.1 Long Term Preservation

National Archives is creating a facsimile of original documents in the form of microfilming and keeping them away under ambient conditions for posterity and for effective disaster management.

Positive Microfilms are prepared from the original negative Microfilms and supplied to scholars and users of Archives for consultation, thus avoiding the handling of old and fragile documents.

4.1.1 Preservation Challenges

No standard formats are available for the image formats.

Integration of images with reference media data [digital images of analog microfilm with data base]

Periodical checking of images in CDs and DVDs [whether the Hardware, drives would be available in future?]

Lack of migration policy [will it be 5 years or 7 years?]

Compatibility of software and hardware [update of software for compatibility with new hardware]

New technological advancements

Non-standardized technical activities

Shortage of skilled personnel

Management policy

4.1.2 Preservation policy

The mandate of Archives is to preserve and conserve archival material housed in its custody. For Paper Records there are set guidelines and laid down official policy. Electronic Records, from the archival perspective, is in the stage of infancy. For its application in the field of conservation of records it requires urgent and immediate action. National Archives of India has now initiated a policy for preservation of records in Archives and Libraries. The same is under consideration by experts group.

Suggestions

An Office Procedure Manual for Electronic Records both born digital (email, etc) and converted to digital mode in line with the stipulations incorporated in 'Manual of Office Procedure" should be drawn within a specific time frame.

The Government of India has initiated in 1999 a "security microfilming programme of 'A' category files which are still in custody of creating agencies.

In a similar manner till such time a final decision is taken regard to the retention schedule/ appraisal/ transfer/ access of Electronic Records (both born digital and converted to digital), the Ministry of Information Technology should initiate in collaboration with National Archives of India, a 'DIGITAL RECORDS PRESERVATION PROGRAMME" by creating a corpus of experts to visit various Ministries/Departments and take 'back up' of electronic records in appropriate medium which would subsequently be migrated to latest technology in keeping pace with the changing technology in this particular field. (This would also take care of Disaster Management of Electronic Records towards preserving the cultural heritage for posterity.

National Archives being the nodal agency for upkeep of paper and other formats of records as per the mandate of Public Record Act and Public Record Rules, therefore, Ministry of Information Technology should take the initiative to develop National Archives as a hub for preserving the digital information for the benefit of scholars and users of Archives for the retrieval of information.

By conducting regular awareness programme for the benefit of personnel in the Ministries and Departments and thus apprising them of the latest developments in the field of information technology.

5. MIGRATION

Migration ensures the periodic transfer of digital information from one software/ hardware configuration to another available medium in vogue, or from one generation of computer technology to a subsequent generation of information technology. If there is a potent danger for loss of valuable information during migration then institutions would have to determine what should be a total acceptable quantum of loss (risk management) primarily with access in mind, and serve as watchword towards evolution of a definite policy.

The various aspects of digital preservation, which should be strictly followed is: -

1. How is preservation of digital object different than preserving an analog object?
2. How does preservation depend on access?
3. What are the essential attributes of a digital object that should be preserved?
4. How should you manage preservation of the enormous volume of digital objects created by governments and transferred in disk/CD formats to archives for preservation?

The professionals in archives are often faced with complex and numerous related issues confronting their established tradition, practice and ethics. Digital preservation represents a significant new paradigm for archivists for archival collections and their management.

In view of the electronic records being generated with speed there appears a need to evolve preservation policy at the national level on priority basis. To meet this challenge Archives should evolve and develop policy guidelines, for these techniques.

5.1 Objectives

The objectives of a digital preservation policy, alongwith its purpose and scope, are consistent with those objectives typically identified for traditional media but are broadened or extended to account for a new medium and process, and work out the priorities for their use and access. Digital preservation revolves around the choice of interim storage media, the life expectancy of a digital imaging system, and the expectation to migrate the digital files to future systems while ensuring full functionality and the integrity of the original digital system

It is a universally accepted fact that digital technology has lowered the cost of data storage and increased the quantum of research data available, but digital technologies alone are not sufficient to make this data useful. In its first place, stored data has to be inherently "useful" for answering a worthwhile research query. Secondly just because information is less costly for storage does not necessarily make it more accessible, especially for research consultation.

Archivists' concern the world over, relate to the authenticity and integrity of data are implied and, therefore, deserve special attention. In archives, maintaining the value and importance of digital holdings ensures that the information accessed by the user is the same information that was initially acquired and described and which still relays its context and structure. As stated earlier, the care of manipulation of digital data, either intentional or accidental, leads to deliberate alteration or deletion of information without leaving any trace or evidence.

The principle objective for archives should be the document management of the digital record over a period of time in order to guarantee the reliability of its content and meaning. An objective of digital preservation policy should be to clarify and extend the application of ideas to a new medium or to introduce new concepts keeping in view the necessary adherence to the history, provenance of archives, etc.

Digital preservation policy should include and highlight the timely provision of advice or provide support to the areas of collections, acquisition and operations. This is crucial in the digital world as technological obsolescence is a greatest challenge to the acquisition and intellectual control activity of the concerned archival institution. The institution should bring the object under "preservation control" (ensuring current and future access) with the technology necessary to access its informational content. This issue becomes critical when the institution has a considerable backlog of programmes relating to acquisitions or preservation due to resource crunch or other underlying factors.

Traditionally a successful track record is necessary to launch any new endeavor. In the case of digital preservation they provide an effective launch pad for the development of a new policy model for the information age, which builds on existing knowledge and incorporates added layers of technological and professional concerns. Thus, traditional preservation policy offers a foundation which represents a mature approach to the preservation of varied collections of cultural heritage and reflects advances in professional ethics and institutional vision.

Established preservation principles relating to respect for the integrity of the original, minimal intervention, reversibility, as well as concepts of preventive preservation, custody and access can only partially support the demands of digital collections.

Traditional preservation principles are so substantially transformed when applied to digital material that in practice they eventually fail to describe the scenario in a meaningful manner. Frequent debates on the nature and relevance of preservation in a digital world especially the role of archivists, librarians, curators, managers, politicians, consumers, manufacturers is necessary for clarifying principles which encompass the entire gamut of preservation of all cultural resources, digital as well as other formats.

Digital information is also being lost to future generation of researchers, especially after purpose of the data is fulfilled; the user has little incentive to protect the data for historical purposes. In India presently when the information is displayed in the public domain, it is not organized for the benefit of historical researchers. Despite all the formal and prescribed mechanisms designed to help researchers to retrieve the more traditional original archival sources, they find that unsystematic informal communication would be absolutely essential for retrieval of digital information which is not kept with historical interests in mind.

While formulating preservation policy for Archives as far as digitization is concerned, the following features should be taken in to account: -

1. Co-operation – national level management of Digital Archives/Repositories – the State Archives have a major role to play in it.

2. Redundancy of digital equipments hardware as well as software

3. International standard inter- portability

The policy should spell out answers and solutions for digital media longevity, rates of deterioration of digital objects, storage conditions and disaster recovery. The retention period of the paper records and electronic records should also be looked into so as to ensure an appropriate retention period of the electronic records is fixed. This could be deliberated in high-level techno-professionals' forum representing the archival fraternity alongwith information technological experts.

The subject of digital preservation is still in the primary domain of the information technology professionals. The acquisition, preservation and management of objects in new medium does pose a challenge to Archives. Besides, there has been an increase in the quantum of information being created digitally in Government. The speeds with which new technologies to preserve information are applied have not kept pace with its creation.

The digital images may not last long due to fast changing technology in hardware medium every 3-4 years and it would be not possible to keep pace with ever changing technology with ever-growing volume of records. Besides, problems would also arise to provide on site technical back up to the record creating agencies as compared to a central place of records storage, i.e. National Archives of India.

To preserve the contents an acceptable approach to digital preservation would have to be found out; otherwise we would be confronted with the problem of data loss, which would greatly affect the concept archives which are continuous series of records. It may be mentioned that recently equipment are being marketed to convert digital images directly to microfilm media, the life expectancy of this microfilm media is more than 500 years. This should be explored to save the digital images.

Going with the times, the National Archives of India started the computerization of reference media of records in the year 1998. A tailor-made software AIMS (Archival Information Management System), keeping in view the nature of records in its custody, was developed for the purpose. Approximately data entry of reference media of 2.4 million records has already been completed and the same is available for being retrieved through the AIMS software, for scholars and users of Archives in the Intranet of the Department.

The AIMS software, is being used in the Department, has been further updated in the latest technology with SQL backend and dot net in the front end. The software has been upgraded in such a manner that the entire record holdings i.e., of Private Archives, Cartography, Oriental Records and Reprography Section reference media available in paper or microfilm form would be data entered. Moreover the software has been customized for integration of the digital images of the microfilms of records so that when a user wants to consult a document the same could be accessed online and a print out taken, if needed. This has facilitated the consultation of a single original by many scholars in tandem in the intranet in National Archives.

It was reported some time back that the number of internet users worldwide have been increasing 200 % every year. The rapid pace of technological development and obsolescence, threatens vast amounts of information of being "stranded" on tradition or old medium as new and better techniques supersede them. Current approaches to the preservation of important digital information rely on techniques, which should be applied within the life span of medium and their dependent technologies (software, hardware). Apathy to look for trusted means and methods of digital preservation would certainly exact a stiff, long-term cultural penalty.

Adapting preservation policy to the digital age is dependent on the coordination of a complex set of layered determinants of a cultural, a technological and a professional nature. This requires an appreciation of the larger cultural context for framing long term institutional policy and a significant knowledge of digital technology systems and trends. This would lead to emulating current preservation practices and principles against a new digital paradigm, and finally it demands new attitudes to the preservation of cultural heritage collections in digital form. However, the financial issues could be deliberated upon once a consensus is evolved on the policy for digital long time preservation. Financial issues are equally important and are likely to be supported by information experts' advice on cost-effectiveness of hardware and software. This measure would definitely ensure smooth migration on the available latest technology during every ten years in a planned and phased manner.

It may not be an exaggeration to state that, yesterday's world was well-documented but the present digital world is not so. The digital documents stored in hard drives, magnetic tape or digital discs - have storage life of less than 20 years. This necessitates frequent update or migration of data to save contents thus stored therein.

Archivists and data storage professionals are concerned with the longevity the information stored in the hard disk, CDs, magnetic tape and other like mediums.

Recently we have discussions in the net that the so called above problems are going to be solved by the introduction of new Millennial DiscTM* and Millennial WriterTM products which are currently being developed at Millennia, Inc. The technology first developed at Brigham Young University in Provo, Utah, Millennia, Inc. claims that they have since created the first digital disc that safely stores data for 1,000 years. They also claim that the Millennia Disc is backwards compatible, which means once it is created, it can be read like any other recordable DVD, and it requires no special drive to read the contents. However it requires a special writer to write the contents on the Disc This Disc and Writer are scheduled for testing in 2009's first quarter. [*http://millenniata.com]

It seems that in future all the information generated by the Government of India would be in digital form. Emergence of "Digital Government" is a new trend. Manuscripts and Printed Library material are being scanned and stored in digital format. Each nation now dares to create and preserve records or evidences of its activities in a digital format. Any work schedule is unavoidable. The Government of India has been working out viable agendas for introducing e-governance in all Ministries and Departments. Office procedure automation software for storing electronic records in digital form for consultation by posterity is in the offing. The Government of India should ensure that all Acts, Rules and Circulars are available in a user-friendly IT version. The Government of India will surely emerge as one of the most efficient and transparent e-government areas of the world in not too distant future.

6. REFERENCES

[1] Adapting PRESERVATION Policy IN ARCHIVES, TO THE DIGITAL AGE, JOHN A. GRACE